Probabilistic Ontologies for Multi-INT Fusion

Kathryn Blackmond Laskey, Paulo C. G. Costa
Center of Excellence in C^4I
George Mason University, MS 4B5
Fairfax, VA 22030-4444
[klaskey, pcosta]@gmu.edu

Terry Janssen
Lockheed Martin Corporation, IS&GS/GSS
13560 Dulles Technology Drive
Herndon, VA 20171
terry.janssen@lmco.com

Abstract

Systems are increasingly required to fuse data from geographically dispersed, heterogeneous information sources to produce up-to-date, mission-relevant results. These products focus not only on traditional military forces and systems, but to an increasing degree also on non-traditional combatants and their social networks. Successful multi-INT fusion requires that the constituent systems interoperate not just at the level of syntax and formats, but also at the level of semantics. Ontologies are vital enablers for semantic interoperability. Because uncertainty is a fundamental aspect of multi-INT fusion, lack of support for uncertainty is a major limitation of current-generation ontology formalisms. Probabilistic OWL (PR-OWL) extends the OWL Web Ontology Language to enable the construction of probabilistic ontologies. Ontologies constructed in PR-OWL can represent complex patterns of evidential relationships among uncertain hypotheses. Recently, a system for specifying and reasoning with PR-OWL ontologies has been released in alpha version. This paper describes the PR-OWL ontology language, the probabilistic logic on which it is based, and the reasoning system implementation. A hypothetical case study in the counterterrorism domain is illustrates the capabilities of PR-OWL.

1. Introduction

Multi-INT fusion is a critical technology for the next generation of military and intelligence systems. As connectivity and bandwidth increases, commanders and analysts are deluged with ever-greater volumes of data from geographically dispersed, heterogeneous information sources. In today's military engagements, fusion products must focus not only on traditional military forces and systems, but also on non-traditional combatants and their social networks. Successful multi-INT fusion requires that the constituent systems interoperate not just at the level of syntax and formats, but also at the level of semantics. That is, interoperating systems should interpret terminology in a consistent way; or if not, appropriate translations must be established between vocabularies used by different systems. Techniques for making semantic information explicit and computationally accessible are key to effective exploitation of data from diverse sources. Shared formal semantics enables systems with different internal representations to exchange information, and provides a means to enforce business rules such as access controls for security.

When heterogeneous systems are required to interoperate in an open world, vocabularies that were developed for individual stand-alone applications break down. Ontologies provide shared representations of the entities and relationships characterizing a domain, into which vocabularies of legacy systems can be mapped. However, a major limitation of traditional ontology formalisms is the lack of consistent support for uncertainty. Because uncertainty is a fundamental aspect of multi-INT fusion, this is a serious deficiency. Current ontology formalisms provide no principled means to ensure semantic consistency with respect to issues of uncertainty or data quality.

Probabilistic ontologies [1] augment standard ontologies with probabilistic information about the domain. Probabilistic OWL (PR-OWL) extends the OWL Web Ontology Language to enable the construction of probabilistic ontologies. PR-OWL is based on Multi-Entity Bayesian Networks (MEBN), a first-order probabilistic logic that combines the representational power of first-order logic (FOL) and Bayesian Networks (BN) [2]. Ontologies constructed in PR-OWL can represent complex patterns of evidential relationships among uncertain hypotheses. Recently, a system for specifying and reasoning with PR-OWL ontologies has been released in alpha version [3, 4]. This system, called UnBBayes-MEBN,

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
21 MAY 2008	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Probabilistic Ontologies for Multi-INT Fusion	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Center of Excellence in C4I George Mason University, MS 4B5 Fairfax, VA 22030-4444	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
AFCEA-GMU C4I Center Symposium "Critical Issues In C4I" 20-21 May 2008, George Mason University, Fairfax, Virginia Campus, The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**22**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

provides a graphical user interface for defining entities, attributes, and probabilistic relationships, defining instances, entering evidence, and entering queries. It also includes a reasoning system for performing Bayesian inference to calculate responses to probabilistic queries.

The following section describes the PR-OWL ontology language and the MEBN logic on which it is based. Section 3 describes the UnBBayes system for entering and reasoning with PR-OWL probabilistic ontologies. Section 5 illustrates the capabilities of PR-OWL with a hypothetical case study in the counterterrorism domain.

2. Probabilistic Ontologies

Initial attempts to represent uncertainty in ontology languages tend to begin with constructs for attaching probabilities as attributes of entities. This approach is clearly inadequate, in that it fails to account for structural features such as conditional dependence (or independence), double counting of influence on multiply connected graphs, and context-specific independence. Many researchers have pointed out the importance of structural information in probabilistic models (e.g. [5, 6, 7]), and it is well known that some questions about evidence can be answered entirely in structural terms (e.g., [6], page 271). For instance, Shafer ([8], pages 5-9) stated that probability is more about structure than it is about numbers.

This is particularly true in domains such as intelligence analysis and Human Intelligence (HUMINT), which rely on complex chains of argument with many interacting uncertain hypotheses, in which subtle features of an argument may augment or diminish its force [6]. Structural information also plays a major role in the way evidence collected from multiple sensors with different degrees of reliability and trust is evaluated. In many cases, different aspects of the same piece of information have to be analyzed and weighed based on incomplete knowledge about the source. Structural information is a key asset to provide an in-depth analysis of what each piece of knowledge means in the overall context of an evidential chain. Special-purpose stand-alone systems may not explicitly represent many of these subtle structural features, leaving them as implicit assumptions underlying the algorithmic processing performed by the system. However, when systems interoperate, it is essential to represent explicitly the assumptions underlying the processing, and to share information about the context of reasoning, to enable the consuming system to properly assess credibility of the information and its import within the overall context of

reasoning. That is, systems must share not only conclusions, but semantic information about how those conclusions were reached and the conditions under which the conclusions are valid. This requires semantic interoperability.

State-of-the-art systems are increasingly adopting ontologies as a means to ensure formal semantic support for knowledge sharing. Uncertainty is becoming recognized as an important aspect to be represented and used in reasoning. A common mistake is to provide support for uncertainty representation by simply annotating ontologies with numerical probabilities. This is a weak approach that leads to fragile intelligence systems, as too much information is lost to the lack of a representational scheme that can capture structural nuances of the probabilistic information. Clearly, more than mere annotation is needed. Indeed, there is a need for a new category of ontologies.

Definition 1 (from [1]): A *probabilistic ontology* is an explicit, formal knowledge representation that expresses knowledge about a domain of application. This includes:

1a. Types of entities that exist in the domain;
1b. Properties of those entities;
1c. Relationships among entities;
1d. Processes and events that happen with those entities;
1e. Statistical regularities that characterize the domain;
1f. Inconclusive, ambiguous, incomplete, unreliable, and dissonant knowledge related to entities of the domain; and
1g. Uncertainty about all the above forms of knowledge;

where the term entity refers to any concept (real or fictitious, concrete or abstract) that can be described and reasoned about within the domain of application. ∎

Probabilistic ontologies are used for the purpose of comprehensively describing knowledge about a domain, along with its associated uncertainty, in a principled, structured and sharable way. Ideally, this knowledge should be represented in a format that can be read and processed by a computer. Probabilistic ontologies also expand the possibilities of standard ontologies by introducing the requirement of a proper representation of the statistical regularities in a domain, and uncertain evidence about entities in a domain of application.

Another aspect that must be emphasized when devising data sharing schemes for intelligence systems is the level of expressivity of a representation

formalism. In other words, any representational scheme that attempts to convey all the details and idiosyncrasies of a complex domain must be highly expressive. Although tractability requirements often motivate restrictions on the ability of reasoning engines to process highly expressive representations, if ontologies are to be general-purpose repositories of shared knowledge, then restrictions on reasoners should not dictate what it is possible to say about a domain. PR-OWL, which is used in this paper as the language for building probabilistic ontologies, can achieve the required level of expressivity because it is based on a First-Order Bayesian Logic that represents probability distributions over interpretations of arbitrary first-order domain theories [2].

MEBN is a first-order Bayesian logic that integrates classical first-order logic with probability theory. Classical first-order logic (FOL) is by far the most commonly used, studied and implemented logical system, serving as the logical basis for most current-generation AI systems and ontology languages. MEBN logic provides a logical foundation for extending the capability of ontology languages to include a logically coherent representation for uncertainty. Because a MEBN theory represents a coherent probability distribution, Bayes Theorem provides a mathematical foundation for learning and inference that reduces to classical logic in the case of certain knowledge (i.e., all probabilities are zero or one).

MEBN represents the world as comprised of entities that have attributes and are related to other entities. Knowledge about the attributes of entities and their relationships to each other is represented as a collection of MEBN fragments (MFrags) organized into MEBN Theories (MTheories). An MFrag represents a small, repeatable piece of knowledge about the probabilistic relationships among a set of interrelated hypotheses about attributes of or relationships among entities of given types. The generic knowledge represented by the MFrag can be instantiated repeatedly on different entities of the allowable types, thus composing complex argument structures from repeated sub-structures. An example of this is shown below in the case study.

Specifically, an MFrag contains context, input, and resident random variables (RVs), a fragment graph and local distributions. The RVs represent uncertain hypotheses; the fragment graph represents dependency relationships the RVs; and the local distributions provide quantitative information about the strength of the relationships encoded by the fragment graph. Together, the fragment graph and the local distributions define conditional probability distributions for instances of the resident random variables (RVs), conditional on the values of instances

of their parents in the fragment graphs, and given the context constraints. Distributions for the input and context RVs are defined in other MFrags. Context nodes represent conditions assumed for definition of the local distributions.

A collection of MFrags that satisfies certain consistency constraints implicitly defines a joint probability distribution on instances of its random variables. Such a collection of MFrags is called an MTheory. MEBN semantics integrates the standard model-theoretic semantics of classical first-order logic with random variables as formalized in mathematical statistics. Specifically, a theory in first-order logic defines a set of possible worlds; and any world in which all the axioms of the theory are satisfied is called a model of the axioms. Beyond ruling out worlds inconsistent with the axioms, classical logic cannot say anything about relative plausibility of the possible worlds. A first-order Bayesian logic such as MEBN can grade the possible worlds according to plausibility. Thus, from a given set of axioms, first-order logic can do no more than assert that an assertion is proven, disproven, or neither proven nor disproven. As with FOL, MEBN logic assigns probability zero to assertions that can be disproven from the axioms of an MTheory, and probability one to assertions that can be proven. However, MEBN logic can assign probabilities between zero and one to hypotheses that can be neither proven nor disproven.

As a full integration of first-order logic and probability, MEBN provides: (1) a means of expressing a globally consistent joint distribution over models of any consistent, finitely axiomatizable FOL theory; (2) a proof theory capable of identifying in-consistent theories in finitely many steps and converging to correct responses to probabilistic queries; and (3) a built in mechanism for adding sequences of new axioms and refining theories in the light of observations. Thus, even the most complex situations can be represented in MEBN, provided they can represented in FOL. Furthermore, because MEBN is a first order Bayesian logic, its use as the underlying semantics of PR-OWL not only guarantees a formal mathematical foundation for a probabilistic extension to the OWL language (PR-OWL), but also ensures that the advantages of Bayesian Inference (e.g. natural "Occam's Razor", support for learning from data, etc.) will accrue to PR-OWL probabilistic ontologies. A comprehensive explanation of MEBN logic is not on the scope of this paper, but the interested reader is directed to [2].

PR-OWL was developed as an extension enabling OWL ontologies to represent complex Bayesian models in a way that is flexible enough to be used by diverse Bayesian probabilistic tools (e.g. Netica,

Hugin, Quiddity*Suite, JavaBayes, etc.) based on different probabilistic technologies (e.g. PRMs, BNs, etc.). More specifically, PR-OWL is an upper ontology for probabilistic systems that can be used as a framework for developing probabilistic ontologies (as defined above) that are expressive enough to represent even the most complex probabilistic models. DaConta et al. define an upper ontology as a set of integrated ontologies that characterizes a set of basic commonsense knowledge notions ([9], page 230). In PR-OWL, these basic commonsense notions are related to representing uncertainty in a principled way using OWL syntax (itself a specialization of XML syntax), providing a set of constructs that can be employed to build probabilistic ontologies.

Figure 1 shows the main concepts involved in defining an MTheory in PR-OWL.
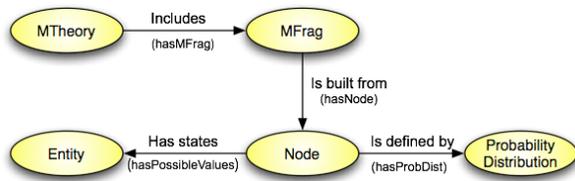
Figure 1 – Main Elements of PR-OWL

In the diagram, ellipses represent general classes while arrows represent the main relationships between these classes. A probabilistic ontology (PO) has to have at least one individual of class MTheory, which is basically a label linking a group of MFrags that collectively form a valid MTheory. In actual PR-OWL syntax, that link is expressed via the object property hasMFrag (which is the inverse of object property isMFragIn). Individuals of class MFrag are comprised of nodes, which can be resident, input, or context nodes (not shown in the picture). Each individual of class Node is a random variable RV and thus has a mutually exclusive, collectively exhaustive set of possible states. In PR-OWL, the object property hasPossibleValues links each node with its possible states, which are individuals of class Entity. Finally, random variables (represented by the class Nodes in PR-OWL) have unconditional or conditional probability distributions, which are represented by class ProbabilityDistribution and linked to its respective nodes via the object property hasProbDist. Figure 2 depicts the main elements of the PR-OWL language, its subclasses, and the secondary elements necessary for representing an MTheory. The relations necessary to express the complex structure of MEBN probabilistic models using the OWL syntax are also depicted. In addition to [1] the prospective reader will

find more information on the PR-OWL language at http://www.pr-owl.org.
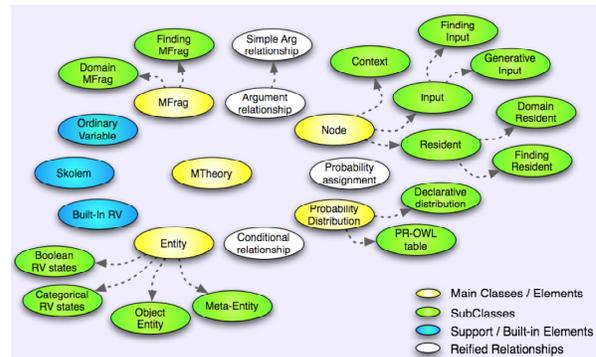
Figure 2 – PR-OWL Elements

The first step towards building a probabilistic ontology in compliance with our Definition 1 is to import into any OWL editor an OWL file containing the PR-OWL classes, subclasses, and properties (one is available at http://www.pr-owl.org/pr-owl.owl). After importing the PR-OWL definitions, the next step in ontology design is to construct domain-specific concepts, using the PR-OWL definitions to represent uncertainty about their attributes and relationships. Using this procedure, an ontology engineer is not only able to build a coherent generative MTheory and other probabilistic ontology elements, but also make it compatible with other ontologies that use PR-OWL concepts.

3. A Reasoner for Bayesian Ontologies

At its current stage of development, PR-OWL contains only the basic representation elements that provide a means of representing any MEBN theory. Such a representation could be used by a Bayesian tool (acting as a probabilistic ontology reasoner) to perform inferences to answer queries and/or to learn from newly incoming evidence via Bayesian learning.

However, building MFrags in a traditional ontology editor is a manual, error prone, and tedious process. Avoiding errors or inconsistencies requires deep knowledge of the logic and of the data structures of PR-OWL, since the user would have to know all technical terms such as hasPossibleValues, isNodeFrom, isResidentNodeIn, etc. Furthermore, reasoning with a PR-OWL ontology requires creating instances of the random variables need to respond to a given query, assembling them into a Bayesian network, and entering that Bayesian network into a software application that can perform the desired inference. This too is a tedious, manual, error-prone process. Ideally, much of this work could be automated by a software

application designed to enforce the consistency of a MEBN model and to respond correctly to queries.

The development of UnBBayes-MEBN, an open source, Java-based application that is currently in alpha phase (public release March 08), is an important step towards this objective, as it provides both a GUI for building probabilistic ontologies and a reasoner based on the PR-OWL/MEBN framework.

UnBBayes-MEBN was designed to allow building probabilistic ontologies in an intuitive way without having to rely on a deep knowledge of the PR-OWL specification. Figure 3 shows a snapshot of the UnBBayes-MEBN user interface. In the figure, a click on the "R" icon and another click anywhere in the editing panel will create a resident node, for which a description can be inserted in the text area at the lower left part of the screen. Clicking on the arrow icon would allow one to graphically define the probabilistic relations of that resident node with other nodes, as much as it would be done in current Bayesian packages such as Hugin™. All those actions would result in the software creating the respective PR-OWL tags (syntactic elements that denote particular parts of a PR-OWL ontology) in the background.
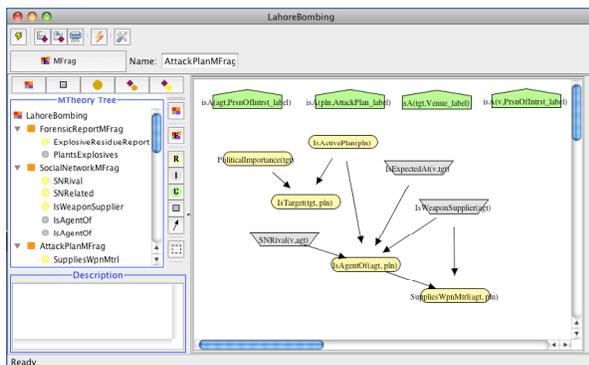
Figure 3 – The UnBBayes-MEBN GUI

Probabilistic Ontologies in UnBBayes-MEBN are saved in PR-OWL format (*.owl file), while application-specific data is stored in a text file with the *.ubf extension. Support for MEBN input/output operations is provided via the Protégé-OWL API[1], which is based on the class JenaOWLModel. By using a common API, UnBBayes-MEBN ensures that MTheories created using its GUI can be opened and edited in popular ontology editor Protégé[2] (and vice-versa). This compatibility is important because it ensures that files created in UnBBayes-MEBN can be opened and edited not only in Protégé, but also in any OWL-compliant application (although these

---

[1] http://protege.stanford.edu/plugins/owl/api/index.html

[2] http://protege.stanford.edu

applications will not be able to understand the ontology's probabilistic characteristics). In addition, ontologies that have already been defined using an OWL-compliant editor can be extended to the PR-OWL format in a quick and direct way. All that is needed is to open the OWL file in UnBBayes-MEBN, create an MTheory for the ontology, and save the result.

UnBBayes-MEBN provides not only a GUI for building probabilistic ontologies, but also a probabilistic reasoner that allows for plausible inferences to the knowledge base (KB) using Bayes Theorem as evidence accrues. Currently, only a restricted class of queries has been implemented, but future releases will include the ability to perform multiple queries at the same time.

When a query is submitted, the knowledge base is searched for information to answer the query. If the available information does not suffice, then the KB and the generative MTheory are used to construct a BN to answer the query. This process is called Situation Specific Bayesian Network (SSBN) construction.

In the current implementation, a query consists of a single random variable (RV) instance, which is not allowed to have any evidence below it. The following procedure takes a node name and a list of entity instances as arguments. It is called initially with the query node and its arguments.

PROCEDURE SSBN-CNSTR(NODE,ENTITY-LIST)
(i) For the RV instance NODE(ENTITY-LIST), search for evidence in the KB. If there is a finding for this given entry, finish.
(ii) Search for the resident node that has the name NODE and get its MFrag. Once NODE(OV-LIST) is found, verify if the type of ENTITY-LIST is the same as OV-LIST (where OV-LIST is the list of ordinary variable arguments for NODE in its home MFrag).
(iii) Verify in the KB which context nodes refer to the OVs in OV-LIST, replacing each OV by the appropriate instance in ENTITY-LIST. If any context variable is false, mark the MFrag to use the default distribution.
(iv) If the truth-value of the context node in (iii) is not determined, make it a parent of NODE.
(v) For each parent of NODE, identify any instance of the parent that can be constructed by replacing the OVs by the known entities (contained in the query or KB), and has not yet been added to the SSBN. For each such parent instance, call procedure SSBN-CNSTR for the parent node and its arguments.
(vi) Create the NODE's CPT.
(vii) Finish.

This algorithm is easily enhanced to allow multiple query nodes and evidence below query nodes. These enhancements are currently under development.

A few performance issues had to be considered in implementing UnBBayes-MEBN. Depending on the complexity of the domain, the algorithm may reach a context node that cannot be immediately evaluated. This happens when all ordinary variables in the parent set of a resident random variable term do not appear in the resident term itself. In this case, there may be an arbitrary, possibly infinite number of instances of a parent for any given instance of the child. In this case, the local distribution for a random variable must specify how to combine influences from all relevant instances of its parents.

However, especially in complex formulas this may have a strong impact in the performance of the algorithm, so the designed solution involves asking the user for more information. In the current implementation, if one does not provide such information the algorithm will just halt.

Another design option was to restrict memory usage in a way that a possible memory overload triggers a warning to the user and stops the algorithm. In step (iii), a design optimization over the general SSBN algorithm in [2], only the necessary context nodes for a given MFrag are evaluated, in contrast with the original solution of revising all the context nodes for that MFrag.

Although the implementation addressed other optimization issues, for the sake of conciseness only the most relevant are listed here. UNBBayes-MEBN is a work in progress that is still in alpha status, but it already provides a major contribution to the development of probabilistic ontologies. Its basic functionality was enough to support our work in designing a case study employing POs as a knowledge sharing enabler.

4. Case Study: Attack in Lahore

To illustrate the capabilities of PR-OWL to represent the kinds of multi-INT fusion problems faced by today's net-centric systems, we consider a hypothetical counter-terrorism case study. Our simple illustrative scenario concerns an attempted attack on a high-profile meeting in Pakistan that is detected and prevented through collaboration between two intelligence analysts and interoperation of diverse fusion systems. Although the scenario is hypothetical, it illustrates the role of semantic technology and probabilistic reasoning in enabling a successful intervention to prevent a terrorist plot from succeeding.

The analysts. Intelligence analyst IA1 has been assigned the task of compiling and maintaining social networks of persons-of-interest in Pakistan. Over time, he has developed a social network that includes a known arms dealer (AD) in Islamabad and his associates. Meanwhile, intelligence analyst IA2 has been tasked with compiling and maintaining an intelligence profile of the city of Lahore. In this role, IA2 has access to all intelligence reports associated with people, events, communications, etc within his area of responsibility (AOR).

The meeting. At present, IA2 is aware of, and is monitoring, a conference of six Tribal Leaders (TL1 – TL6) that is occurring in Lahore. This is a high-profile meeting that is receiving heavy coverage by news agencies all over the world, and is therefore of concern as a potential terrorist target.

The arrest. At the Lahore airport, a canine unit has detected explosive residue on a Lahore resident (P) attempting to leave the city. Upon receiving this report, IA2 declares P a person-of-interest. This declaration initiates an automatic action to add P to the scope of IA1's social network, and to alert IA1 to report any significant results concerning P coming from the social network analysis. IA1's analysis uncovers a third-order relationship between P and AD: P's brother, BP, has the same religious advisor, C, as AD.

Figure 4 shows a set of MFrags that could be used to support the above analysis. These MFrags are shown as screenshots from the UNBBayes-MEBN system. The MFrags involve reasoning about entities of different types and the relationships among them. In an operational analyst support system, the PR-OWL ontology that represented the uncertain aspects of this problem would import existing upper ontologies and domain ontologies. For this illustration, we constructed a simple, stand-alone PR-OWL ontology.

Plan Agent and Target. This MFrag represents basic information about attacks using explosives. The context random variables, drawn as pentagons at the top of the MFrag, represent logical conditions assumed to hold when the probability distributions are assigned. In this case, the context random variables state that *pln* represents an attack plan, *agt* and *v* represent persons-of-interest, and *tgt* represents a venue that might be targeted by the attack. In our simple example, we take a venue to denote a localized space-time region that might be the focus of an attack. The MFrag contains two input random variables, whose distributions are defined in other MFrags. These are shown as trapezoids in the figure. They represent whether *agt* is a weapons supplier and whether *v* and *agt* are rivals in the social network. Root nodes in the MFrag are random variables representing whether the plan is active, the political importance of the target, and
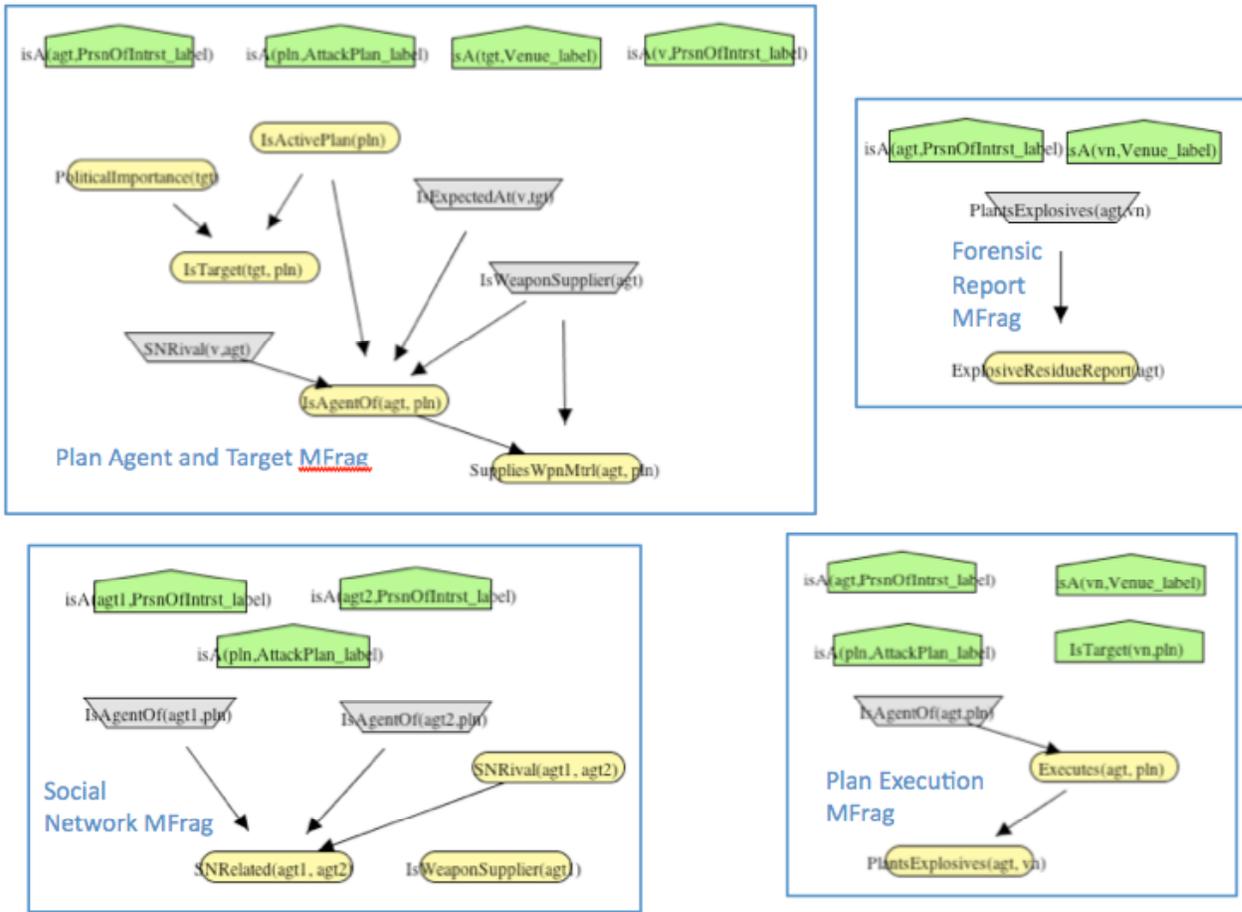
Figure 4: MFrags for First SSBN

whether a rival of *agt* is expected to be present at the venue. Whether the venue is targeted depends on whether the plan is active (if the plan is not active, then no venue is targeted by the plan), and the political importance of the venue (important venues are more likely to be targeted. Whether *agt* is an agent of the plan, i.e., is actively involved in bringing it about, depends on whether the plan is active, whether a rival of *agt* is expected to be at the venue (agents may try to target their rivals), and whether the agent is a weapons supplier (weapons suppliers are more likely to be agents in attacks using explosives). Finally, whether *agt* plays the role of supplying weapons depends on whether *agt* is an agent of the plan and whether *agt* is a weapons supplier.

Social Network. This MFrag represents the actors and their relationships. Its context variables state that *agt1* and *agt2* are persons-of-interest and *pln* is an attack plan. It represents the knowledge that two agents of the same plan are likely to be related in the social network. It also represents probabilities that two

persons-of-interest are rivals and that a person-of-interest is a weapons dealer.

Plan Execution. This MFrag represents the knowledge that an agent of a plan may execute the plan, and one of the activities a plan executor might perform is to plant explosives at the targeted venue.

Forensic Report. This MFrag represents the possibility that an individual who plants explosives may be apprehended and explosive residues detected.

Of course, the model described here is highly simplified – its purpose is to illustrate the capabilities of the language and not to provide a sophisticated representation of terrorist attacks. Our ability to represent the problem is limited by the inability of the alpha implementation of UNBBayes-MEBN to represent subtypes. We expect this limitation to be removed in future versions.

We could not use UNBBayes-MEBN to construct a situation-specific Bayesian network (SSBN) for IA2's analysis problem, because the current version is limited to the special case of a single query node with no evidence below the query node. Our model does not
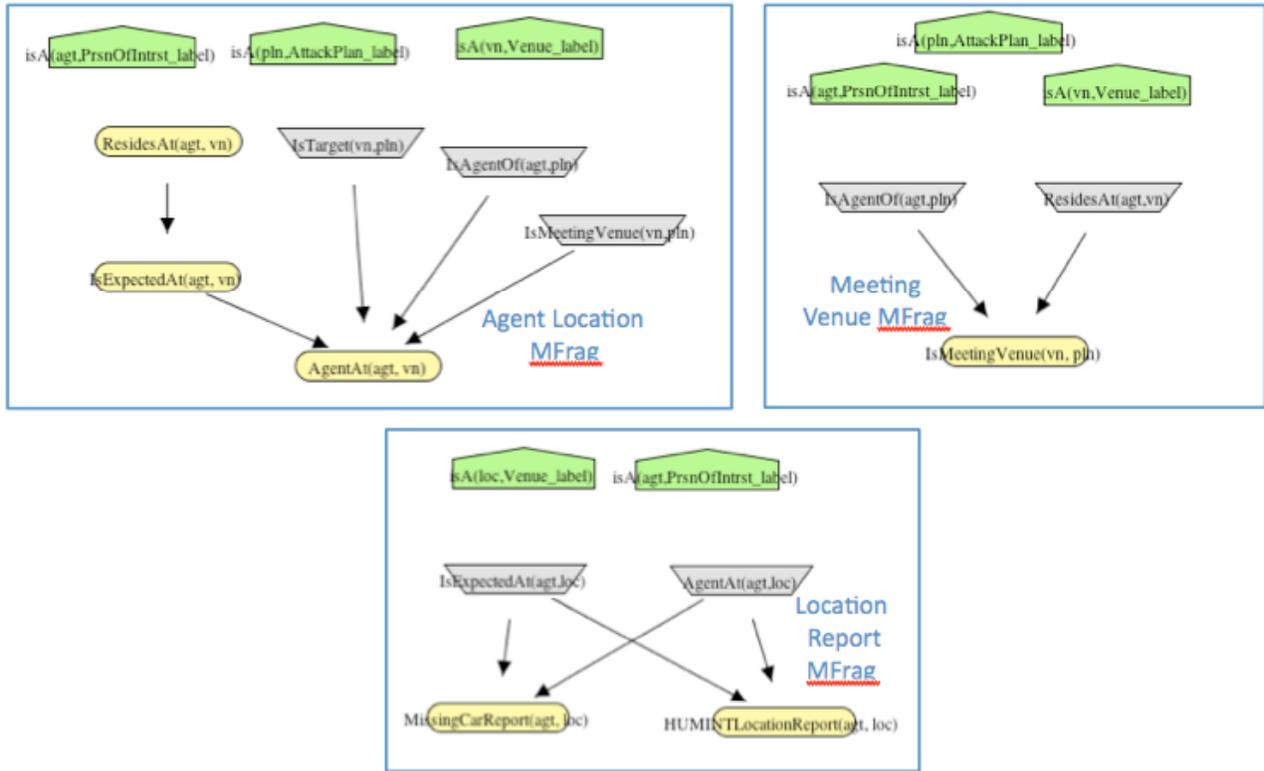
**Figure 5: MFrags Added to Create Second SSBN**

meet that limitation. Nevertheless, we did construct a SSBN by hand for this problem. To do this, we first defined instances of the relevant entities: *P*, *C*, and *AD* (persons of interest), *Conf* (a venue), and *ConfAtk* (an attack plan). The query of concern is whether the conference is targeted. This is represented by the RV instance *IsTarget*(*Conf*, *ConfAttack*). This is an instance of the generic RV *IsTarget*(*v*, *pln*) in the Plan Agent and Target MFrag, in which *Conf* has been substituted for the ordinary variable *v* and *ConfAtk* has been substituted for the ordinary variable *pln*. Evidence random variable instances *ExplosiveResidueReport*(*P*), *SNRelated*(*P*, *AD*), *SNRelated*(*C*, *AD*), *IsWeapon-Supplier*(*AD*), and *PoliticalImportance*(*Conf*) are also created to represent the information that explosive residues were found on P, that P and AD are related through the social network analysis, that C and AD are related through the social network analysis, that AD is an arms dealer, and that the conference has high political importance. SSBN construction begins with these RV instances, identifies any additional RV instances needed to compute a query response, instantiates them, and uses the fragment graphs to compose them into a Bayesian network. After declaring the evidence, a standard belief propagation algorithm is used to compute a query response.

The prior probability of an arbitrary venue being targeted for an attack was set at 0.02%. For an event of high political importance such as the conference in question, the probability is 0.3%. After incorporating the information that AD is a weapons dealer who is related in the social network to both C and P, and that explosive residues were detected on P, the probability that the conference has been targeted for attack becomes about 10%.

As part of his continuing analysis, IA2 has been monitoring the system for current intelligence information related to the conference. A query for the current locations of TL1 through TL6 reveals a HUMINT report that TL6 was seen in Karachi five hours ago. A query for IMINT change detection indicates that a vehicle that was present during the conference is now missing from the conference location. A further analysis of the HUMINT report reveals that TL6 was seen entering the residence of C. Finally, a query to the social network system reveals that TL6 and TL5 are bitter rivals.

Figure 5 shows a set of MFrags that can be instantiated to incorporate this new information.

Agent Location. This MFrag represents the knowledge that an individual who is expected at a
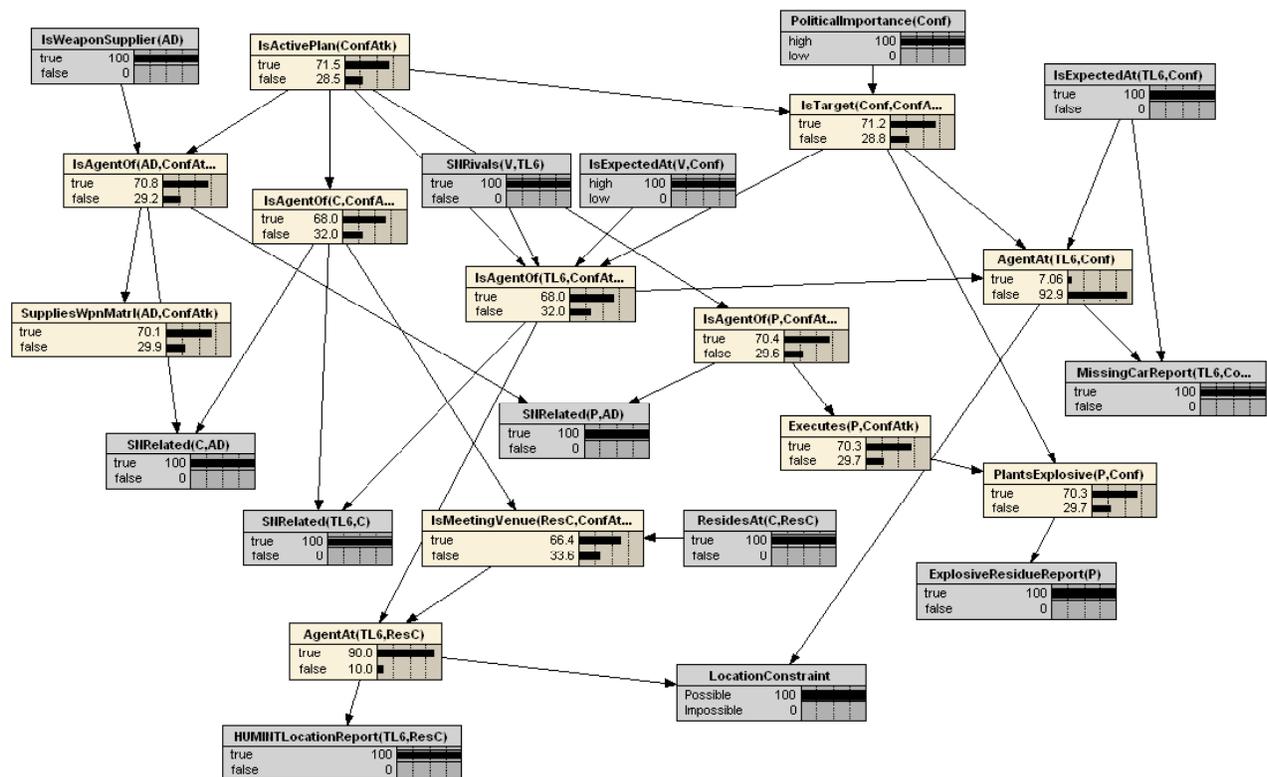
Figure 6: Situation-Specific Bayesian Network for Conference Bombing

venue is likely to be at the venue unless the individual is an agent of a plan that targets the venue.

Coordination MFrag. This MFrag represents the knowledge that agents of a plan may meet to coordinate their plan.

HUMINT Report MFrag. This MFrag represents the information .

After constructing the situation-specific Bayesian network and adding the evidence that TL6 was reported to be entering C's residence, that C's residence was a location other than the conference, that TL6 and C are related in the social network (inferred by logical reasoning from the visit to C's residence), that TL6's car was missing from the conference, and that TL6 and TL5 are rivals, the probability has increased to about 71% that the conference has been targeted for an attack.

Figure 6 shows the SSBN constructed by hand using the Netica® Bayesian network software package. Comparing this SSBN with the MFrags, we see that its random variables are instances of the random variables from the MFrags, obtained by substituting problem-specific entity instances for the ordinary variables of the MFrags. We are currently extending the SSBN

construction algorithm in UnBBayes-MEBN to be capable of constructing the SSBN for this problem.

This problem requires bringing knowledge to bear about events in space and time, how agents use objects such cars, social interactions among agents, and other sophisticated kinds of reasoning. Many of these reasoning patterns are reusable across a wide variety of problems. Examples include the knowledge that individuals may meet with each other to coordinate joint activities, and that they use cars for transportation. In an operational system, these kinds of reasoning would make use of existing ontologies. PR-OWL allows the user of such an ontology to add probabilistic information to represent relationships that fall short of certainty.

To conclude our case study, after using PR-OWL and Bayesian reasoning to explore the implications of the evidence, IA2 appreciates the significance of the combined Multi-INT data, and issues an Actionable Intelligence Report to interdict the possible terrorist attack.

5. Conclusion

Exactly how ontologies should work with probabilities is still an open research issue. The Intelligence Analysis of the knowledge sharing use case presented in this work has shown how probabilistic ontologies can be used to address that issue. UnBBayes-MEBN, which was used to support the use case, is still in alpha phase and should see various improvements in the near future. This system is a promising environment for building probabilistic ontologies to support knowledge sharing in open world environments.

Ontologies provide the "semantic glue" to enable knowledge sharing among diverse systems cooperating in data rich domains such as Intelligence Analysis, but fail to provide adequate support for uncertainty, an ubiquitous characteristic of open world environments. Effective multi-INT fusion requires uncertainty management to be effective, and recent advances in research on probabilistic ontologies have the potential to integrate uncertainty management smoothly with semantic technology.

The case study presented in this work has shown that such research, albeit in its infancy, can help to support interoperability among Intelligence systems in an open environment, addressing issues of fusing multiple sources of noisy information into a coherent overall situation picture.

6. Acknowledgements

10. References

[1] Costa, P. C. G. 2005. Bayesian Semantics for the Semantic Web. PhD Diss. Department of Systems Engineering and Operations Research, George Mason University. 315p, July 2005, Fairfax, VA, USA.

[2] Laskey, K.B., MEBN: A Language for First-Order Bayesian Knowledge Bases, Artificial Intelligence, 172(2-3), 2007.

[3] Carvalho, R. N., Santos, L. L., Ladeira, M., and Costa, P. C. G. 2007. A Tool for Plausible Reasoning in the Semantic Web using MEBN. In *Proceedings of the Seventh International Conference on Intelligent Systems Design and Applications*, 381-386. IEEE Press.

[4] Costa, P. C. G.; Ladeira, M.; Carvalho, R. N.; Laskey, K. B.; Santos, L. L.; and Matsumoto, S. 2008. A First-Order Bayesian Tool for Probabilistic Ontologies. In Proceedings of the 21st Florida FLAIRS UNBBayes-MEBN

[5] Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Mateo, CA, USA: Morgan Kaufmann Publishers.

[6] Schum, D.A. 1994. Evidential Foundations of Probabilistic Reasoning, John Wiley & Sons, Inc., New York, NY.

[7] Kadane, J. B.; and Schum, D. A. 1996. A Probabilistic Analysis of the Sacco and Vanzetti Evidence. New York, NY, USA: John Wiley & Sons.

[8] Shafer, G. 1988. Combining AI and OR. University of Kansas School of Business, Working Paper No. 195. April.

[9] DaConta, M. C.; Obrst, L. O.; and Smith, K. T. 2003. The Sematic Web: A Guide to the Future of Xml, Web Services, and Knowledge Management. Indianapolis, IN, USA: Wiley Publishing, Inc.

Probabilistic Ontologies for Multi-INT Fusion

Kathryn Blackmond Laskey
George Mason University

Paulo C. G. da Costa
George Mason University

Terry Janssen
Lockheed Martin Corporation

Critical Issues in C4I
AFCEA/GMU Symposium
May, 2008

Intelligence Analysis and Uncertainty

- Intelligence information comes from reports subject to many kinds of uncertainty
  - Noise in sensors
  - Incorrect, incomplete, deceptive human intelligence
  - Lack of understanding of cause and effect mechanisms in the world
  - etc.
- Effective intelligence analysis requires combining uncertain reports from multiple sources to form a coherent picture of a situation

2
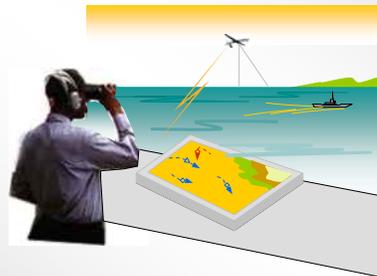
Vision: A Net-Centric World

- Autonomous software agents interoperate seamlessly
- Each agent has timely access to mission-critical information
- Agents are not overloaded with unnecessary information
- Information is properly synchronized and up-to-date
- Data from disparate sources is fused into mission-relevant knowledge
- Multi-level security permits needed access while preventing non-authorized use
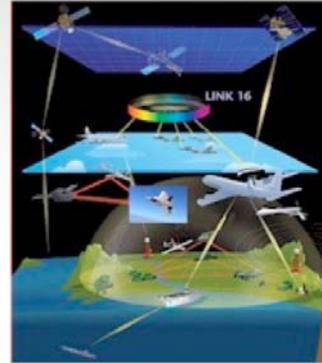
Web Services: Enabling Interoperability
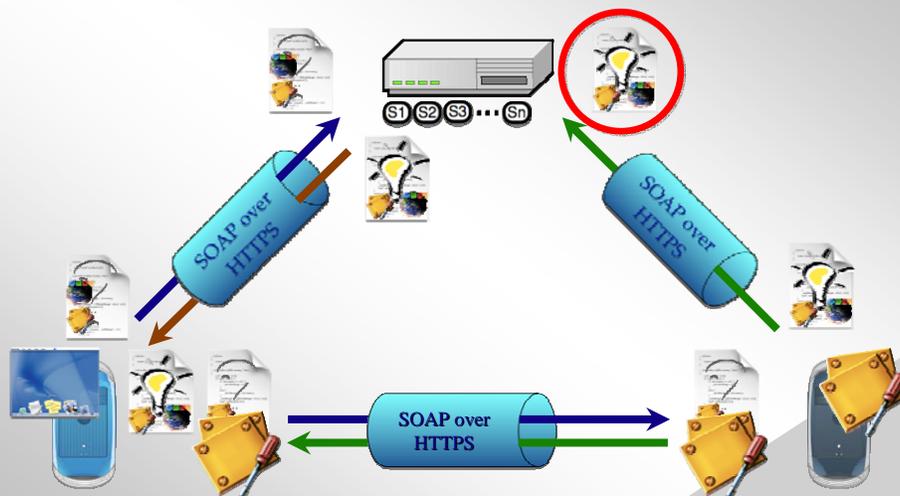
The P-F-B Triangle
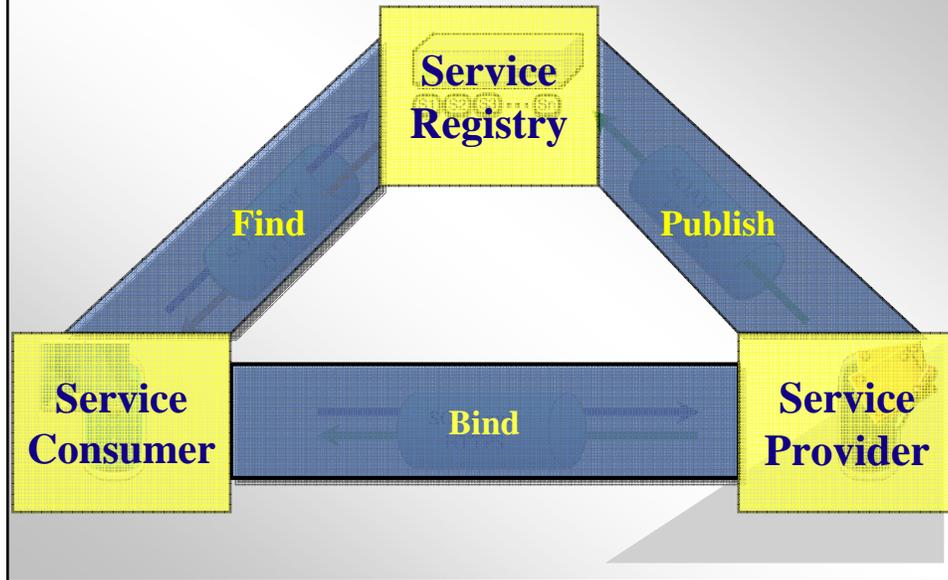
Why Semantics?

Syntax
- *Syntax*: rules of formation for a data type
- *Syntactic interoperability*: applications can process each other's data formats
- *Example*: 3.2 is a legal floating point number

Semantics
- *Semantics*: the meaning of expressions
- *Semantic interoperability*: applications interpret data in the same way
- *Example*: Diagnostic benchmarks were run on 3.2 GHz processor
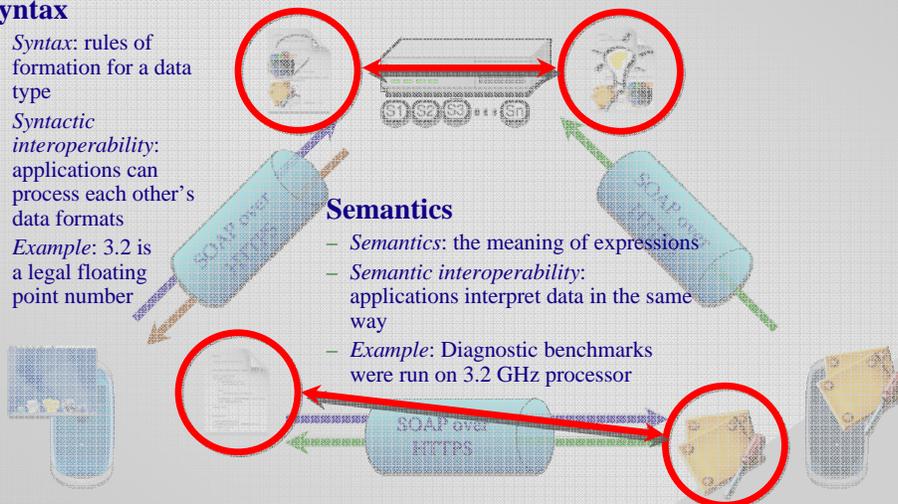
Semantic interoperability is a much stronger requirement than type consistency

Semantics in Net-Centric Services

- Semantics in stovepipe systems are in the mind of the human
  - Natural language documentation
  - Data structures embedded in code
- Net-centric systems require formal, machine-interpretable semantics
- Semantic information in service descriptions enables consumers and providers to have a common understanding of:
  - What does the service do?
  - What inputs does it require and what results does it produce?
  - What are conditions (constraints/policies) for use?
  - How to invoke it? (Address & WSDL description)

7

Uncertainty and Ontologies

- Semantically aware systems are essential to multi-INT fusion.

- Ontologies are a means to semantic awareness

  - Explicit, formal representation of entities and relationships in a domain of application

- Representing and reasoning with uncertainty is essential

- But...

  TRADITIONAL ONTOLOGICAL ENGINEERING METHODS PROVIDE NO SUPPORT FOR REPRESENTING AND REASONING WITH UNCERTAINTY IN A PRINCIPLED WAY

8

Issues

- How should uncertainty be represented in semantically aware systems?
- Should ontologies provide a means to express uncertainty?
- How should ontologies and probabilistic reasoning systems work together?

Probability and Ontology

- Much of our domain knowledge is statistical
- Reasoner can use statistical regularities to:
  - Classify instances (Bayesian classifiers)
  - Infer attributes of instances
- Multi-INT fusion systems must exchange more than just conclusions
  - Uncertainties
  - Pedigrees
  - Sources and credibility information
- Representing statistical information in ontologies supports interoperability

PR-OWL: A Language for Expressing Probabilistic Ontologies

- Extends W3C recommended OWL ontology language
- Based on expressive probabilistic logic
- Represents probabilistic knowledge in XML-compliant format.
- Open-source, freely available solution for representing knowledge and associated uncertainty in a principled manner.
- Reasoner under development at University of Brasilia
  - Alpha version released March, 2008 on SourceForge

(Costa, 2005)

11

PR-OWL classes
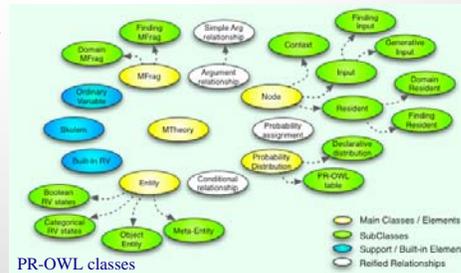
Case Study: Bombing in Lahore

12

Background   Information Set #1

- Roles:
  - IA1& IA2 – Intelligence Analysts #1&2
  - AD – Known Arms Dealer in Isamabad
  - TL1 – TL6 – Tribal Leaders
- Background information:
  - IA1 maintains social networks of persons-of-interest in Pakistan; has created a SN around AD
  - IA2 has access to all intel reports associated with Lahore
  - IA2 is currently monitoring a conference of six tribal leaders (TL1-TL6) which is occurring in Lahore

13

- New Roles:
  - P – Person arrested in Lahore
  - B_P – The brother of P
  - C – Religious Advisor, Cleric
- New Evidence:
  - Canine unit detects explosive residue on P attempting to leave city
  - P is declared a Person-Of-Interest and added to IA1's SN
  - SNA reveals:
    - B_P is the brother of P
    - C is the religious advisor of B_P
    - C is also the religious advisor of AD
 - IA1 alerts IA2 of relationship

# UNBBayes-MEBN

Multi-Entity Bayesian Networks

- Synthesis of Bayesian networks and first-order logic
  - MEBN is to Bayesian networks as algebra is to arithmetic
- MEBN fragments (MFrags) represent probabilistic relationships among small set of related uncertain hypotheses
- Compose into MEBN Theories (MTheories)
  - Collection of MFrags that satisfies consistency constraints
  - represents probability distribution over model structures of associated first-order logic theory
- Use situation-specific BN (SSBN) to reason over instances

15

Lahore Incident MFrags I
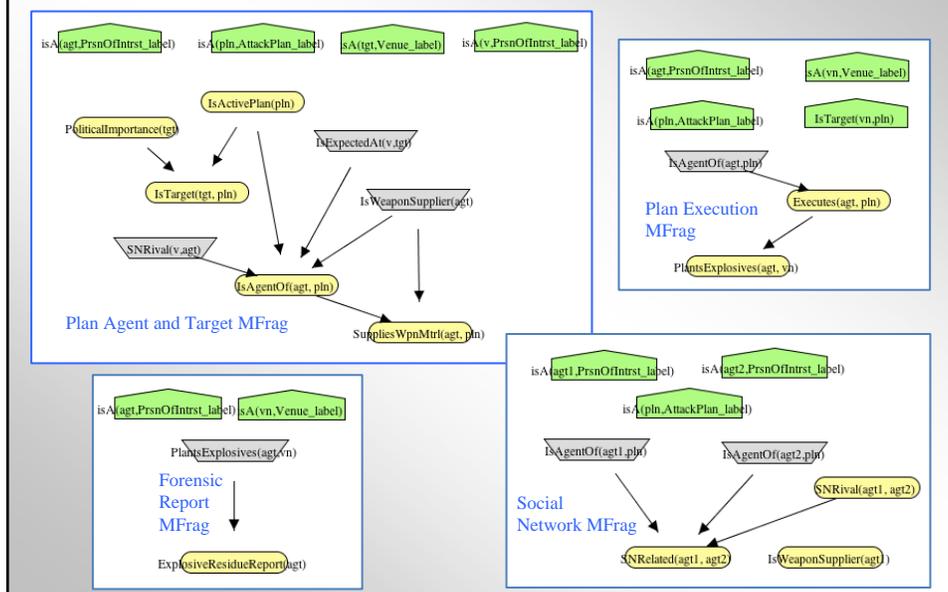
SSBN Before Adding Evidence
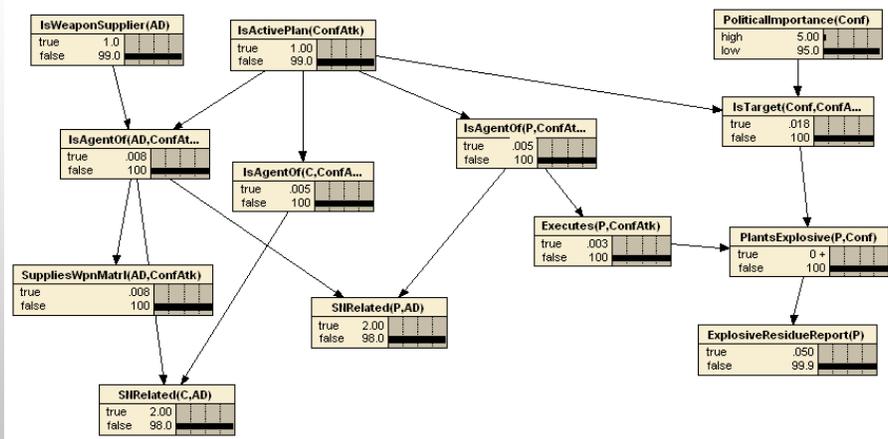
SSBN With Evidence
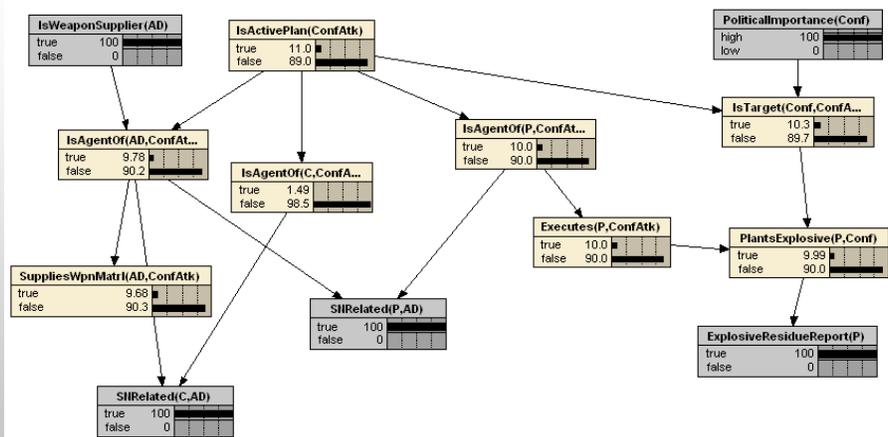
- Probability of attack on conference has increased from 0.2% to about 10%

Information Set #2:

- New evidence:
  - HUMINT report that TL6 is seen in Karachi entering residence of C
    - C is religious advisor of AD and brother of arrested person P
  - IMINT change detection reports missing car from site of conference

19
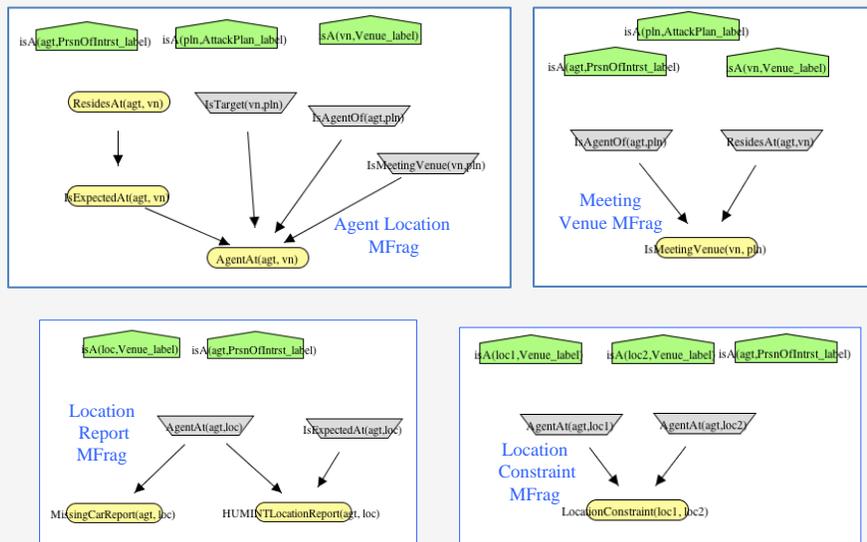
Lahore Incident MFrags II

Augmented SSBN

Probability of attack has increased to 71%
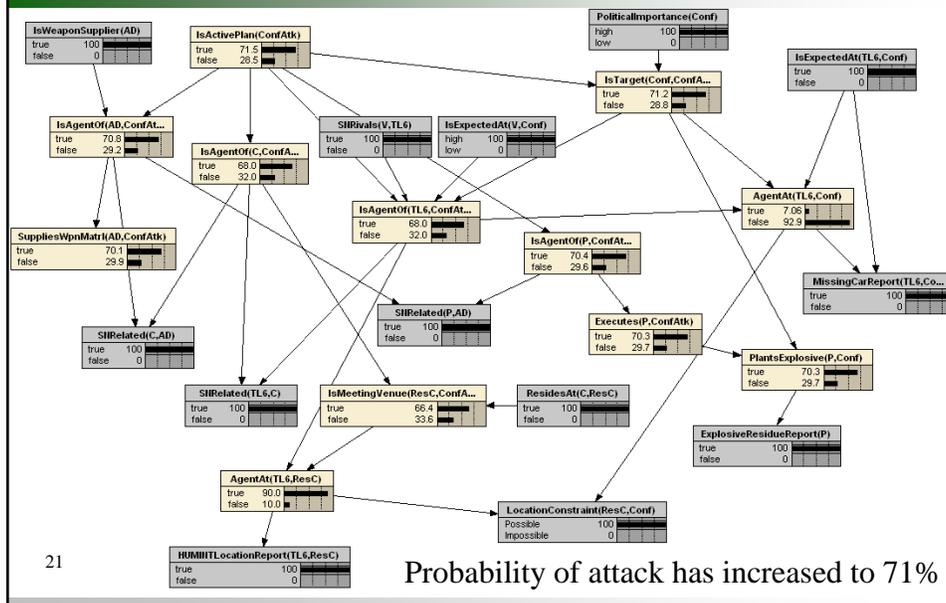
Comments

- This case study illustrates aspects of reasoning that are needed to achieve effective multi-INT fusion
- Enablers for automated support for this kind of reasoning:
  - General knowledge of logical constraints on properties of and relationships among entities of different types
  - General knowledge about likelihoods for properties of and relationships among entities of different types
  - Computationally efficient reasoner for building SSBN from instance-specific reports
- Enablers for multi-INT fusion
  - Shared vocabularies for interchanging all of the above types of information

In Conclusion…

- Uncertainty is ubiquitous in intelligence analysis
- Effective multi-INT fusion requires uncertainty management
- Uncertainty management must work smoothly with semantic technology
- PR-OWL extends OWL ontology language to represent probabilistic information

Thank You!